Estimation in semiparametric time series regression[*]

Jia Chen, Jiti Gao[†] and Degui Li

In this paper, we consider a semiparametric time series regression model and establish a set of identification conditions such that the model under discussion is both identifiable and estimable. We estimate the parameters in the model by using the method of moment and the nonlinear function by using the local linear method, and establish the asymptotic distributions for the proposed estimators. We then discuss how to estimate a sequence of local departure functions nonparametrically when the null hypothesis is rejected and establish some related asymptotic theory. Both the simulation study and the empirical application are also provided to illustrate the finite sample behavior of the proposed models and methods.

AMS 2000 subject classifications: 62F12, 62G05, 62G20.
Keywords and phrases: Asymptotic distribution, Departure function, Local linear method, Semiparametric modelling.

1. INTRODUCTION

Various estimation and specification testing problems in time series analysis have been proposed and discussed extensively in recent years. Interest focuses on general nonparametric and semiparametric time series models under stationarity assumption. Recent studies include [1, 3, 4, 8, 11, 12] as well as the references therein. In the semiparametric case, interest is on the estimation and specification testing in a semiparametric time series model when there are at least two different time series involved. In both theory and practice, there is some need to establish the mathematical relationship between one time series and another, and then discuss both the estimation and specification testing in such a model. When the same time series variable is fully involved in both the parametric and nonparametric components of a semiparametric time series regression model, to the best of our knowledge, the issue of how to identify and estimate the model has not been addressed.

This paper starts with a semiparametric time series model of the form

$$(1.1) \qquad Y_t = V_t^\tau \beta + \Delta(V_t) + e_t, \quad t = 1, 2, \ldots, n,$$

where $\{V_t\}$ is a stationary sequence of d-dimensional random vectors, β is a vector of unknown parameters, $\Delta(\cdot)$ is an unknown function defined on R^d, $\{e_t\}$ is a sequence of independent and identically distributed (i.i.d.) errors, and n is the number of observations. This paper focuses on the case of $1 \le d \le 3$. In the case of $d \ge 4$, to avoid "the curse of dimensionality", one may need to approximate $\Delta(\cdot)$ by a partial sum of univariate functions in a similar fashion to Section 2.3 of [4].

Model (1.1) has different types of motivations and applications to the conventional semiparametric time series model of the form $Y_t = U_t^\tau \beta + \Delta(V_t) + e_t$, in which U_t and V_t are two different stationary time series such that $\Sigma = E[(U_t - E[U_t|V_t])(U_t - E[U_t|V_t])^\tau]$ is a positive definite matrix. In model (1.1), the linear component in many cases plays the leading role while the nonparametric component behaves like a type of unknown departure from the classic linear model. Since such departure is usually unknown, it is not unreasonable to treat $\Delta(\cdot)$ as a nonparametrically unknown function. In the process of estimating both β and $\Delta(\cdot)$ in model (1.1) consistently, existing methods, as discussed in the literature for the partially linear case in [4, 8] for example, are not valid because $\Sigma = E[(V_t - E[V_t|V_t])(V_t - E[V_t|V_t])^\tau] = 0$. In Section 2 below, we consider a more general semiparametric time series model than model (1.1) and propose using a nonlinear least squares (LS) estimation method to deal with the estimation of the unknown parameter and function involved. Recently, a closely-related paper [2] proposes a unified family of parametrically-guided nonparametric estimation schemes, which combines the merits of both the parametric and nonparametric methods and incorporates the prior information.

In Section 2, we also consider an extension of model (1.1) to cover the case where an extended form of (1.1) becomes a semiparametric model as an alternative involved in the hypotheses:

$$(1.2) \qquad \begin{aligned} H_0 &: \ E[Y_t|V_t = v] = v^\tau \beta_0 \\ \text{versus} \quad H_1 &: \ E[Y_t|V_t = v] = v^\tau \beta_1 + \Delta_n(v), \end{aligned}$$

where β_0 (or β_1) is the true value of the parameter β under H_0 (or H_1), and $\{\Delta_n(v)\}$ is a sequence of nonparametrically unknown departure functions. Interest in the literature mainly focuses on constructing a test for H_0. To the best of our knowledge, existing literature does not provide us with any guidance about how to specify an alternative form and then consistently estimate $\Delta_n(\cdot)$ when H_0 is rejected. The paper [5] suggests using a semiparametric estimation method in the practical implementation of an optimal bandwidth selection method when a kernel-based test is used. Section 2.3 below systematically discusses how to identify and then estimate both β_1 and $\Delta_n(\cdot)$ consistently.

In both model (1.1) and the semiparametric model involved in the alternative hypothesis (1.2), we allow $\{V_t\}$ to have a deterministic trend component. As a consequence, both models are more applicable to deal with the case where the nonstationarity of $\{V_t\}$ is caused by a deterministic trend function. Such cases include consumer price indices and global temperature data. The rest of the paper is organized as follows. Section 2 discusses both the identification and estimation issues. Section 3 mentions some extensions. A simulation study and real data analysis are given in Section 4. Some concluding remarks are given in Section 5. All mathematical proofs are given in Appendix A.

2. IDENTIFICATION AND ESTIMATION

In this section, we first give some sufficient conditions to ensure that the models under discussion are identifiable, and then construct parametric and nonparametric estimation methods for the parameter and the nonlinear function, respectively. In addition, we also establish asymptotic distributions for the proposed estimators.

2.1 Nonlinear LS estimation method

Consider a semiparametric nonlinear time series model of the form

$$(2.1) \qquad Y_t = g(V_t, \theta_1) + \Delta(V_t) + e_t, \quad t = 1, 2, \ldots, n,$$

where $g(\cdot, \theta_1)$ is a known link function indexed by an unknown parameter vector $\theta_1 \in \Theta \subset R^p$ ($p \geq 1$), $\{e_t\}$ is a sequence of i.i.d. random errors, $\Delta(\cdot)$ is an unknown function defined on R^d, and $\{V_t\}$ is generated by

$$(2.2) \qquad V_t = H\left(\frac{t}{n}\right) + u_t,$$

in which $H(\cdot)$ is a d-dimensional vector of unknown functions defined on R and $\{u_t\}$ is a sequence of i.i.d. random errors.

It is easy to see that model (2.1) includes model (1.1) as a particular case. $\{V_t\}$ defined as above allows for the existence of deterministic trends in the regressors. To present the main ideas and make this paper more concise, we consider the case of $d = 1$ in Section 2 and Appendix A. Section 3 discusses how to deal with the case of $d \geq 2$.

As discussed in Section 1, there are various motivations for us to consider a semiparametric time series model of the form (2.1). In the analysis of economic and financial data, one may motivate the proposal of model (2.1) by considering a general parametric nonlinear model of the form

$$(2.3) \qquad Y_t = g(V_t, \theta_1) + \varepsilon_t,$$

where the error process $\{\varepsilon_t\}$ is endogenously correlated with $\{V_t\}$ through an additive model of the form $\varepsilon_t = \Delta(V_t) + e_t$. In such a case, $\{V_t\}$ and $\{\varepsilon_t\}$ are likely to be dependent on each other.

Before we discuss the identifiability and estimability of θ_1 and $\Delta(\cdot)$ in model (2.1), we introduce the following conditions.

A1 (i) $\{u_t\}$ is an i.i.d. sequence, and its density function $p_u(\cdot)$ is continuous.

(ii) $\{e_t\}$ is an i.i.d. sequence with $0 < \sigma_e^2 := E(e_1^2) < \infty$. In addition, $P\{E(e_t|u_t) = 0\} = 1$ and the joint density function $p_{e,u}(\cdot, \cdot)$ of e_t and u_t is continuous.

A2 (i) The nonlinear regression function $g(v, \theta)$ is twice differentiable with respect to θ, and both $\Delta(\cdot)$ and $H(\cdot)$ are continuous.

(ii) Denoting the partial derivative of $g(v, \theta)$ with respect to θ by $\dot{g}(v, \theta) = \frac{\partial g(v, \theta)}{\partial \theta}$, then

$$(2.4)$$
$$\Gamma(\theta) := \int_0^1 \left[\int \Delta(v)\, \dot{g}(v, \theta) p_u(v - H(r)) dv \right] dr = 0$$

for all $\theta \in \Theta$, and

$$(2.5)$$
$$\int_0^1 \left\{ \int [g(v, \theta_1) - g(v, \theta)]\, \dot{g}(v, \theta) p_u(v - H(r)) dv \right\} dr \neq 0$$

uniformly in $\theta \in \Theta(\delta) = \{\theta : \|\theta - \theta_1\| \geq \delta\}$ for any $\delta > 0$.

A3 (i) The matrix

$$\Sigma(\theta_1) := \int_0^1 \left[\int \dot{g}(v, \theta_1) \dot{g}^\tau(v, \theta_1) p_u(v - H(r)) dv \right] dr$$

is positive definite. In addition, $\dot{g}(v, \theta)$ is continuous in v, and the matrix

$$\Sigma_e := \int_0^1 \left[\int \sigma^2(v, r) \dot{g}(v, \theta_1) \dot{g}^\tau(v, \theta_1) p_u(v - H(r)) dv \right] dr$$

is positive definite, where $\sigma^2(v, r) = \int_{-\infty}^{\infty} x^2 p_{e,u}(x, v - H(r)) dx$.

(ii) $\Delta(\cdot)$ is twice continuously differentiable, and the matrix

$$\Sigma_\Delta := \int_0^1 \left[\int \Delta^2(v) \dot{g}(v, \theta_1) \dot{g}^\tau(v, \theta_1) p_u(v - H(r)) dv \right] dr$$

is positive definite.

A4 (i) $K(\cdot)$ is a symmetric and continuous probability density function with $\int K^2(u)du < \infty$ and $\int u^2 K(u)du < \infty$.

  (ii) The bandwidth b satisfies $\lim_{n\to\infty} b = 0$, $\lim_{n\to\infty} nb = \infty$ and $\lim_{n\to\infty} nb^5 < \infty$.

A1 imposes some conditions on the error terms e_t and u_t. The i.i.d. conditions are assumed to simplify the proofs of the main results. In fact, they can be relaxed at the cost of involving more tedious proofs. For example, we can show that the main results still hold under some mixing dependence conditions. A2 is imposed to ensure that θ_1 in (2.1) is identifiable and estimable. In particular, (2.5) in A2(ii) is similar to

$$(2.6) \quad \int_0^1 \left\{ \int [g(v,\theta_1) - g(v,\theta)]^2 p_u(v - H(r))dv \right\} dr > 0$$

uniformly in $\theta \in \Theta(\delta) = \{\theta : \|\theta - \theta_1\| \geq \delta\}$ for any $\delta > 0$. A special case of (2.6) with $H(\cdot) \equiv 0$ is commonly used in the literature (see, for example, [13]). A3 is mainly used for the establishment of the asymptotic theory such as Theorem 2.2 and Theorem 2.3 below. A4 is commonly used in nonparametric kernel estimation (see, for example, [3]).

To estimate both θ_1 and $\Delta(\cdot)$ involved in (2.1), we start with a nonlinear LS estimation method by choosing the true version, θ_1, of θ such that

$$(2.7) \quad \frac{1}{n} \sum_{t=1}^n E[Y_t - g(V_t, \theta_1)]^2$$
$$= \arg\min_\theta \frac{1}{n} \sum_{t=1}^n E[Y_t - g(V_t, \theta)]^2.$$

To ensure that θ_1 is identifiable and estimable, we provide the following proposition; its proof will be given in Appendix A.

Proposition 2.1. *Let* A1 *and* A2 *hold. Then,* θ_1 *is the unique solution of* (2.7).

Based on Proposition 2.1, the sample version of (2.7) suggests using the Method of Moments to estimate θ_1 by

$$(2.8) \quad \widehat{\theta}_1 = \arg\min_\theta \frac{1}{n} \sum_{t=1}^n [Y_t - g(V_t, \theta)]^2.$$

We then estimate $\Delta(\cdot)$ by a local linear estimator of the form

$$(2.9) \quad \widehat{\Delta}(v) = \sum_{t=1}^n W_{nt}(v)[Y_t - g(V_t, \widehat{\theta}_1)],$$

where $\{W_{nt}(v)\}$ is a sequence of weight functions defined by

$$W_{nt}(v) = \frac{K_{v,b}(V_t)}{\sum_{k=1}^n K_{v,b}(V_k)} \quad \text{and} \quad K_{v,b}(V_t) = \frac{1}{b} K_n\left(\frac{V_t - v}{b}\right),$$

with $K_n(\frac{V_t - v}{b}) = K(\frac{V_t - v}{b})[S_{n,2}(v) - (\frac{V_t - v}{b}) S_{n,1}(v)]$ and $S_{n,j}(v) = \frac{1}{nb} \sum_{s=1}^n K(\frac{V_s - v}{b})(\frac{V_s - v}{b})^j$ for $j = 1, 2$, and $K(\cdot)$ and b are the kernel function and bandwidth, respectively.

We then establish an asymptotic theory for $\widehat{\theta}_1$ and $\widehat{\Delta}(\cdot)$ in Theorem 2.2 below.

Theorem 2.2. *Let* (2.1) *and* A1–A3 *hold.*
(i) *Then as* $n \to \infty$,

$$(2.10) \quad \sqrt{n}(\widehat{\theta}_1 - \theta_1) \xrightarrow{d} N(\mathbf{0}, \Sigma^{-1}(\theta_1)(\Sigma_e + \Sigma_\Delta)\Sigma^{-1}(\theta_1)),$$

where $\Sigma(\theta_1)$, Σ_e *and* Σ_Δ *were defined in* A3.
(ii) *If, in addition,* A4 *is satisfied, then as* $n \to \infty$

$$(2.11) \quad \sqrt{nb}(\widehat{\Delta}(v_0) - \Delta(v_0) - b^2 c_{1n}(v_0)) \xrightarrow{d} N(0, \Sigma_1(v_0)),$$

where $c_{1n}(v_0) = \frac{1}{2} \Delta''(v_0) \int u^2 K(u)du + o_P(1)$ *and* $\Sigma_1(v_0) = \frac{\sigma^2(v_0)}{f(v_0)} \int K^2(u)du$, *in which* $\sigma^2(v_0) = \int_0^1 \sigma^2(v_0, r)dr$, $\sigma^2(v_0, r) = \int_{-\infty}^\infty x^2 p_{e,u}(x, v_0 - H(r))dx$ *and* $f(v_0) = \int_0^1 p_u(v_0 - H(r))dr > 0$.

Theorem 2.2(i) shows that the parametric estimator $\widehat{\theta}_1$ has the same root-n rate of convergence as in the parametric linear model. The influence of $\Delta(\cdot)$ and the error process $\{e_t\}$ on the asymptotic distribution is reflected by Σ_Δ and Σ_e in the asymptotic variance matrix. Theorem 2.2(ii) shows that it is achievable to obtain a standard result for the local linear estimator. The detailed proof of Theorem 2.2 will be given in Appendix A.

As discussed in Section 2.2 and then Section 4 below, the proposed estimator $\widehat{\theta}_1$ is more efficient than the conventional weighted least squares estimator.

2.2 Semiparametric weighted LS estimation method

If we follow the literature ([7, 8] for example) by treating model (2.1) as a partially nonlinear model of the form

$$(2.12) \quad Y_t - g(V_t, \theta_1) = \Delta(V_t) + e_t$$

and estimating $\Delta(\cdot)$ by

$$(2.13) \quad \overline{\Delta}(v) = \overline{\Delta}(v, \theta_1) = \sum_{t=1}^n W_{nt}(v)[Y_t - g(V_t, \theta_1)],$$

we will then obtain a semiparametric weighted least squares estimator, $\widetilde{\theta}_1$, defined by

$$(2.14) \quad \widetilde{\theta}_1 = \arg\min_\theta \sum_{t=1}^n [Y_t - g(V_t, \theta) - \overline{\Delta}(V_t, \theta)]^2$$
$$= \arg\min_\theta \sum_{t=1}^n [\widetilde{Y}_t - \widetilde{g}(V_t, \theta)]^2,$$

where $\widetilde{Y}_t = Y_t - \sum_{s=1}^n W_{ns}(V_t)Y_s$ and $\widetilde{g}(V_t, \theta) = g(V_t, \theta) - \sum_{s=1}^n W_{ns}(V_t)g(V_s, \theta)$, in which $W_{ns}(v)$ was defined in (2.9).

Due to the local linear method, one may show, similarly to the proof of Theorem 2.2(ii), that as $n \to \infty$

$$(2.15) \qquad \widetilde{g}(V_t, \theta_1) = (1 + o_P(1))\mathbf{c}_1^\tau g_{20}(V_t, \theta_1)b^2$$
$$\text{and} \quad \widetilde{\Delta}(V_t) = (1 + o_P(1))\Delta''(V_t)b^2,$$

where $g_{20}(v, \theta_1) = \frac{\partial^2 g(v, \theta_1)}{\partial v^2}$, $\widetilde{\Delta}(V_t) = \Delta(V_t) - \sum_{s=1}^{n} W_{ns}(V_t)\Delta(V_s)$, and $\mathbf{c}_1$ is a constant vector.

Analogously to the proof of Theorem 2.2(i), we then have as $n \to \infty$

$$(2.16)$$

$$b^2(\widetilde{\theta}_1 - \theta_1)$$
$$= c_2(1 + o_P(1))\left[\sum_{t=1}^{n} \dot{g}_{20}(V_t, \theta_1)\dot{g}_{20}^\tau(V_t, \theta_1)\right]^{-1}$$
$$\times \sum_{t=1}^{n} \dot{g}_{20}(V_t, \theta_1) \, e_t$$
$$+ c_2 b^2(1 + o_P(1))\left[\sum_{t=1}^{n} \dot{g}_{20}(V_t, \theta_1)\dot{g}_{20}^\tau(V_t, \theta_1)\right]^{-1}$$
$$\times \sum_{t=1}^{n} \dot{g}_{20}(V_t, \theta_1)\Delta''(V_t),$$

where c_2 is a constant, $\dot{g}_{20}(v, \theta_1) = \frac{\partial g_{20}(v, \theta)}{\partial \theta}|_{\theta=\theta_1}$ and we have used $\sum_{t=1}^{n} \dot{g}_{20}(V_t, \theta_1)\widetilde{e}_t = (1 + o_P(1))\sum_{t=1}^{n} \dot{g}_{20}(V_t, \theta_1) \, e_t$, in which $\widetilde{e}_t = e_t - \sum_{s=1}^{n} W_{ns}(V_t)e_s$.

While the first term on the right-hand side of (2.16) is of order $O_P(n^{-1/2})$, the second term may only be of order $O_P(b^2)$, which implies that $\widetilde{\theta}_1$ may be inconsistent as $\widetilde{\theta}_1 - \theta_1 = O_P(1)$. Even in some special cases where $\text{Var}[\sum_{t=1}^{n} \dot{g}_{20}(V_t, \theta_1)\Delta''(V_t)] = O(n)$, we can only show that the rate of convergence of $\widetilde{\theta}_1$ to θ_1 is only proportional to $n^{-1/2}b^{-2}$, which is much slower than the rate of $n^{-1/2}$ for $\widehat{\theta}_1$, because of $b \to 0$. Furthermore, letting $b = cn^{-1/5}$, the rate of convergence of $\widetilde{\theta}_1$ to θ_1 is only proportional to $n^{-1/10}$ for such a special case. This is the main reason we propose using $\widehat{\theta}_1$ rather than $\widetilde{\theta}_1$ in this paper. In general, this is the reasoning why the semiparametric estimation method proposed for the conventional partially nonlinear model $Y_t = g(U_t, \theta_1) + \Delta(V_t) + e_t$, in which U_t and V_t are different sets of regressors, is not directly applicable to the partially nonlinear model (2.1).

2.3 Estimation of local departure functions

We next consider a nonlinear model of the form

$$(2.17) \qquad Y_t = m(V_t) + e_t, \quad t = 1, \ldots, n,$$

where $m(\cdot)$ is a smooth function, $\{e_t\}$ is a sequence of i.i.d. errors and $\{V_t\}$ was defined in (2.2). We are then interested

in estimating a class of local nonparametric departure functions involved in the following alternative hypothesis:

$$(2.18) \qquad H_0: \, m(v) = g(v, \theta_0)$$
$$\text{versus} \quad H_1: m(v) = g(v, \theta_1) + \Delta_n(v),$$

where $\theta_0 \in \Theta$ is the true value of the parameter θ under H_0, $\theta_1 \in \Theta$ and $\{\Delta_n(\cdot)\}$ is a sequence of unknown functions, which are referred to as the local departure functions in this paper.

As discussed in the literature (see, for example, [4, 5, 10, 11]), the choice of this type of semiparametric alternatives is mainly because interest in some cases is to detect whether there is a kind of slight departure from a commonly used parametric form when there is no sufficient evidence to suggest accepting the null hypothesis. Also in such cases, the level of such departure may be unknown and will need to be estimated. To the best of our knowledge, the issue of how to consistently estimate $\Delta_n(\cdot)$ has not been discussed in the literature.

Similarly to condition (2.4) in Proposition 2.1, we assume that for $\theta \in \Theta$,

$$(2.19)$$
$$\Gamma_n(\theta) := \int_0^1 \left[\int \Delta_n(v) \, \dot{g}(v, \theta)p_u(v - H(r))dv\right]dr = O(\delta_n)$$

with $\delta_n \to 0$ as $n \to \infty$. Similarly to (2.8), the resulting estimator of θ_1 is still denoted by $\widehat{\theta}_1$. We then estimate $\Delta_n(v)$ by

$$(2.20) \qquad \widehat{\Delta}_n(v) = \sum_{t=1}^{n} W_{nt}(v)[Y_t - g(V_t, \widehat{\theta}_1)],$$

where $\{W_{nt}(v)\}$ was defined in (2.9).

To establish an asymptotic theory for $\widehat{\theta}_1$ and $\widehat{\Delta}_n(v)$, we need to introduce the following condition:

A5 The local departure function $\Delta_n(v)$ is twice continuously differentiable. In addition, as $n \to \infty$,

$$\delta_{1n} := \int_0^1 \left[\int \|\dot{g}(v, \theta_1)\Delta_n(v)\|^2 p_u(v - H(r))dv\right]dr \to 0.$$

We now establish the asymptotic distributions for the proposed estimators of θ_1 and $\Delta_n(\cdot)$ involved in the alternative hypothesis in (2.18).

Theorem 2.3. *Let* (2.19), A1, A2(i), A3(i) *and* A5 *hold. Suppose that* $\Gamma_n(\theta)$ *is continuous in* θ.
(i) *Then, as* $n \to \infty$,

$$(2.21) \quad \sqrt{n}(\widehat{\theta}_1 - \theta_1 - c_{2n}) \xrightarrow{d} N(\mathbf{0}, \Sigma^{-1}(\theta_1)\Sigma_e\Sigma^{-1}(\theta_1)),$$

where

$$c_{2n} = \int_0^1 \left[\int \Delta_n(v) \ \dot{g}(v, \theta_1) p_u(v - H(r)) dv \right] dr$$
$$+ O_P(n^{-1/2} \delta_{1n})$$
$$= O_P(\delta_n + n^{-1/2} \delta_{1n}).$$

(ii) *If, in addition,* A4 *is satisfied, then as* $n \to \infty$,

$$(2.22) \qquad \sqrt{nb}(\widehat{\Delta}_n(v_0) - \Delta_n(v_0) - b^2 \ c_{3n}(v_0) + \tau_n)$$
$$\xrightarrow{d} N(0, \Sigma_1(v_0)),$$

where $c_{3n}(v_0) = \frac{1}{2} \ \Delta_n''(v_0) \int u^2 K(u) du$ *and* $\tau_n = O_P(n^{-1/2} + \delta_n)$.

Theorem 2.3(i) shows that the parametric estimator of θ_1 still has the root-n rate of convergence. When the dependence of $\Delta_n(\cdot)$ on n is explicitly specified as $\Delta_n(v) = \delta_n \ \Delta(v)$, in which $\Delta(v)$ satisfies A2 and A3 and $\delta_n \to 0$, we have the following corollary; its proof will be given in Appendix A.

Corollary 2.4. *Let* $\Delta_n(v) = \delta_n \ \Delta(v)$ *with* $\delta_n \to 0$. *Suppose that* A1–A3 *are satisfied.*
(i) *Then as* $n \to \infty$,

$$(2.23) \qquad \sqrt{n}(\widehat{\theta}_1 - \theta_1) \xrightarrow{d} N(0, \Sigma^{-1}(\theta_1) \Sigma_e \Sigma^{-1}(\theta_1)).$$

(ii) *If, in addition,* A4 *is satisfied and* $\sqrt{nb^5} \ \delta_n \to 0$, *then as* $n \to \infty$,

$$(2.24) \qquad \sqrt{nb\Delta_n^2(v_0)} \left(\frac{\widehat{\Delta}_n(v_0)}{\Delta_n(v_0)} - 1 \right) \xrightarrow{d} N(0, \Sigma_1(v_0)),$$

where $\Sigma_1(v_0)$ *is defined in Theorem* 2.2(ii).

3. DISCUSSION ON POSSIBLE EXTENSIONS

Section 2 discusses two classes of semiparametric time series models and then establishes asymptotic properties for the proposed estimators for the case of $d = 1$. As discussed in the literature ([1, 4, 11] for example), when d is large, one will need to employ a dimensional-reduction technique to address the issue of the "curse of dimensionality".

One possible approach is to use the following additive model

$$(3.1) \qquad Y_t = g(V_t, \theta_1) + \sum_{j=1}^{d} \Delta_j(V_{t,j}) + e_t,$$

where each $\Delta_j(\cdot)$ is an unknown univariate function defined on R, $V_t = (V_{t,1}, \ldots, V_{t,d})^\tau$, and $\{e_t\}$ is the same as in (2.1). If we assume that A1 and A2 are satisfied with (2.4) replaced by

$$\Gamma^*(\theta)$$
$$= \int_0^1 \left\{ \int \dot{g}(v, \theta) \left[\sum_{j=1}^{d} \Delta_j(v_j) \right] p_u(v - H(r)) dv_1 \ldots dv_d \right\} dr$$
$$= 0$$

for all $\theta \in \Theta$, then the unknown parameter vector θ_1 can still be consistently estimated by $\widehat{\theta}_1$. Function $\Delta(v) = \sum_{j=1}^{d} \Delta_j(v_j)$ can then be estimated as in (2.9) and each of the functions $\Delta_j(\cdot)$ can be estimated by the marginal integration method (see, for example, Section 2.3 of [4]).

Another possible approach is to adopt a semiparametric single-index model of the form

$$(3.2) \qquad Y_t = g(V_t, \theta_1) + \Delta(V_t^\tau \gamma) + e_t,$$

where γ is a vector of unknown parameters. Model (3.2) is an extension of the partially single-index model discussed in [14]. Estimation of θ_1, γ and $\Delta(\cdot)$ is then mainly based on the identifiability and estimability of model (3.2). Establishing the corresponding conditions and results to those given in [14] requires further study and therefore is left for future research.

4. SIMULATION AND EMPIRICAL APPLICATION

In this section, we provide a Monte Carlo simulation study and a real data analysis to illustrate the finite sample performance of the proposed estimation method. We employ the "leave-one-out" cross-validation method to select the bandwidth involved in the estimation of the nonparametric function $\Delta(\cdot)$. A quadratic kernel function of the form $K(u) = \frac{3}{4}(1 - u^2)I(|u| < 1)$ is used throughout this section.

Example 4.1. Consider a pair of regression models

$$(4.1) \qquad Y_t = \theta_1 V_t^2 + \Delta(V_t) + e_t$$
$$\text{and} \quad V_t = H\left(\frac{t}{n}\right) + u_t, \quad t = 1, \ldots, n,$$

where $\theta_1 = 0.8$, $\Delta(v) = 2v$, $H(r) = r - 0.5$, $e_t \overset{\text{i.i.d.}}{\sim} N(0, 0.5^2)$, $u_t \overset{\text{i.i.d.}}{\sim} U(-0.1, 0.1)$, and $\{e_t\}$ and $\{u_t\}$ are independently generated. It is easy to verify that the identification conditions (i.e. A1 and A2) in Proposition 2.1 are satisfied in this example. We generated 1,000 realizations, each consisting of $n = 200$, 500 and 1,000 observations.

Let $\widehat{\theta}_1$ denote the LS estimator of θ_1 introduced in Section 2.1, and $\widetilde{\theta}_1$ denote the semiparametric weighted LS estimator of θ_1 introduced in Section 2.2. The means and standard errors (SEs) of the two estimators of the parameter θ_1 based on 1,000 replications are reported in Table 4.1 with the SEs parenthesized.

Table 4.1. The results for θ_1 in Example 4.1

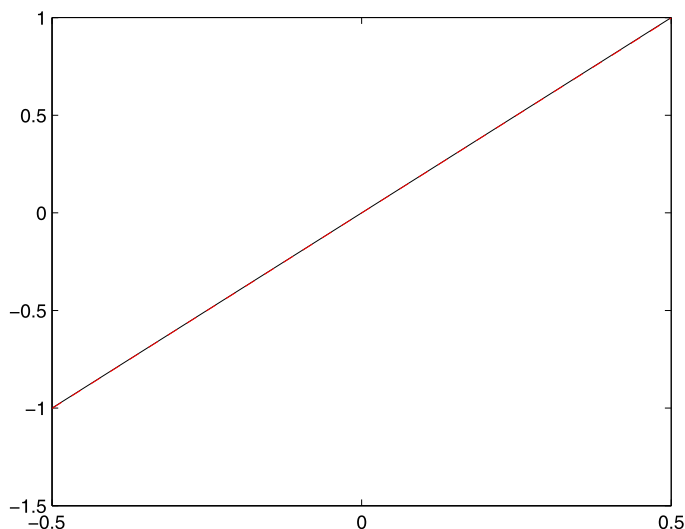| Sample size	$\widehat{\theta}_1$	$\widetilde{\theta}_1$
200	0.9071 (0.3610)	87.2730 (3.8249e + 003)
500	0.8359 (0.2183)	-22.7360 (454.3795)
1,000	0.8196 (0.1611)	3.1732 (363.5797)

Figure 4.1. The true curve $\Delta(v)$ (solid lines) and the mean of $\widehat{\Delta}(v)$ over 1,000 realizations (dashed lines) in Example 4.1 with sample size 1,000.

Table 4.1 shows that the semiparametric weighted LS method, which is commonly used in estimating ordinary partially linear models, fails to provide a good estimate for θ_1. Its estimates deviate far from the true value. This indicates that the semiparametric weighted LS method is not applicable to model (4.1). By contrast, our proposed LS method provides an accurate estimate of θ_1, and its corresponding SE decreases as sample size increases.

The mean estimated curve of $\Delta(\cdot)$ by the LS method at the points $-0.5, -0.4, \ldots, 0.4, 0.5$ is plotted in Figure 4.1, where the solid line represents the true curve and the dashed line represents the estimated curve. As the semiparametric weighted LS method cannot estimate model (4.1) well (as can be seen from Table 4.1), we do not plot the estimated curve resulting from this method. Figure 4.1 shows that the solid line and the dashed line almost coincide, which implies that the local linear estimation method performs very well.

Example 4.2. We next use our model and estimation method to analyze the relationship between the global temperature and Southern Oscillation Index (SOI). The data we used were collected over the period 1866 to 2009 and can be downloaded from the url: http://www.cru.uea.ac.uk/cru/data/. The SOI is defined as the mean sea level pressure difference between Darwin and Tahiti and is known to have a strong association
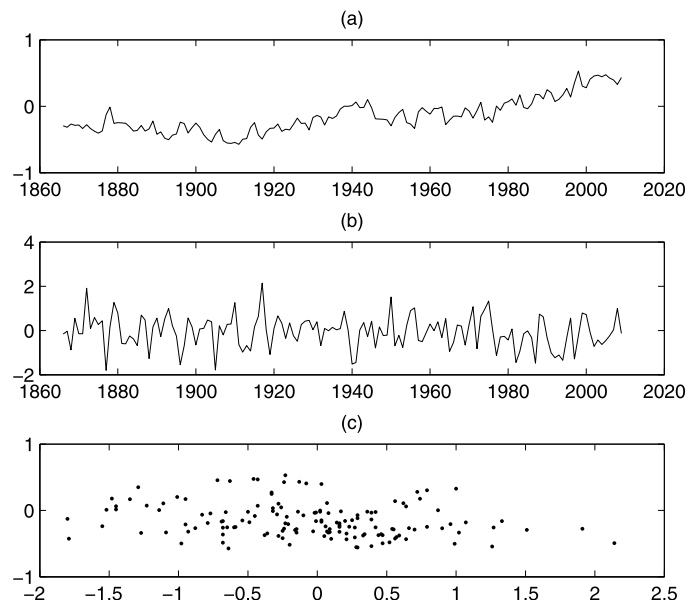

Figure 4.2. (a) The annual mean of the global temperature series (1866–2009), (b) the annual SOI series (1866–2009), and (c) the annual mean of the global temperatures against the annual SOI.

with both global and regional temperature series ([9]). For recent development on the modelling of the relationship between the temperature and SOI, see [6, 15] for references.

Let V_t and Y_t, $1 \le t \le 144$, denote the annual SOI and global temperature of t-th year. Figure 4.2 gives the plots of these two series, as well as the plot of Y_t against V_t. We then fit the data with the following semiparametric model:

$$(4.2) \qquad Y_t = V_t\theta + \Delta(V_t) + e_t,$$

where θ is the parameter and $\{e_t\}$ is the random error. As we allow for the existence of a trend component in V_t, we do not need to make transformations to the raw SOI data, such as filtering or differencing in [15], before fitting the model.

Applying the parametric LS estimation method, we get the estimated value of the parameter as $\widehat{\theta} = -0.0649$ with standard deviation 0.0011. And the nonparametric local linear estimate of the curve $\Delta(\cdot)$, along with its 95% confidence band is given in Figure 4.3, which shows that the values of this estimated curve ranges between -0.1 and -0.4. This indicates that there is a departure from linearity in the relationship between the global temperature and the SOI, and thus a purely linear model might not be able to perfectly capture this relationship. Our semiparametric model (4.2), on the other hand, may provide a better way of analyzing the relationship.

5. CONCLUSIONS

In this paper, we have proposed two classes of semiparametric time series models and then discussed how to identify
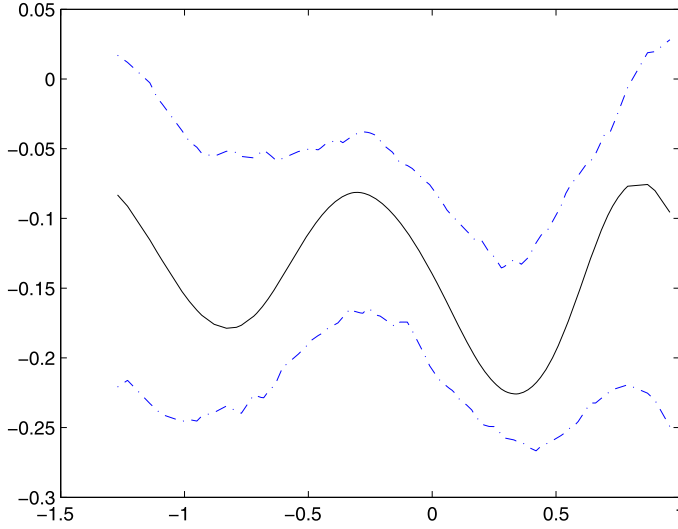
Figure 4.3. The estimated curve of $\Delta(\cdot)$, as well as its 95% confidence band.

and estimate the proposed models. Because of the particular features of the proposed models, existing estimation methods for conventional partially linear models are not applicable. In particular, Section 2.3 has discussed the issue of how to consistently estimate a sequence of nonparametric departure functions in a class of local alternatives involved in a model specification testing problem. Such an estimation procedure may be useful in several aspects, such as studying the power function of a nonparametric test and the choice of a smoothing parameter involved in the nonparametric test. Section 3 has briefly discussed possible extensions. Both simulated and real data examples have been given in Section 4 to show the implementability of the proposed models and methods.

APPENDIX A. PROOFS OF MAIN RESULTS

Proof of Proposition 2.1. Letting

$$\Lambda(\theta) = \frac{1}{n} \sum_{t=1}^{n} E[Y_t - g(V_t, \theta)]^2,$$

we then show that θ_1 is the unique minimizer of $\Lambda(\theta)$. Taking the derivative of $\Lambda(\theta)$ and letting it equal zero, we have

(A.1)
$$\frac{1}{n} \sum_{t=1}^{n} E\{[g(V_t, \theta_1) - g(V_t, \theta) + \Delta(V_t) + e_t]\dot{g}(V_t, \theta)\} = 0.$$

It suffices to show that θ_1 is the unique solution of (A.1). By $P\{E(e_t|u_t) = 0\} = 1$ and (2.4), it is easy to verify that

$$\frac{1}{n} \sum_{t=1}^{n} E[e_t \dot{g}(V_t, \theta)] = 0 \quad \text{and} \quad \frac{1}{n} \sum_{t=1}^{n} E[\Delta(V_t) \, \dot{g}(V_t, \theta)] = 0.$$

Hence, θ_1 is a solution of (A.1). Furthermore, by (2.5), we have

$$\frac{1}{n} \sum_{t=1}^{n} E\{[g(V_t, \theta_1) - g(V_t, \theta)]\dot{g}(V_t, \theta)\} \neq 0$$

uniformly in $\theta \in \Theta(\delta) = \{\theta : \|\theta - \theta_1\| \geq \delta\}$ for any $\delta > 0$. Thus, θ_1 is the unique solution of (A.1). $\qquad\square$

Proof of Theorem 2.2. Observe that

(A.2)
$$\widehat{\theta}_1 - \theta_1 = \left[\sum_{t=1}^{n} \dot{g}(V_t, \theta_1)\dot{g}^{\tau}(V_t, \theta_1) \right]^{-1}$$
$$\times \left\{ \sum_{t=1}^{n} \dot{g}(V_t, \theta_1)[\Delta(V_t) + e_t] \right\}(1 + o_P(1)).$$

By the law of large numbers for i.i.d. sequences, we have

(A.3) $\qquad \frac{1}{n} \sum_{t=1}^{n} \dot{g}(V_t, \theta_1)\dot{g}^{\tau}(V_t, \theta_1) = \Sigma(\theta_1) + o_P(1).$

Hence, to prove Theorem 2.2(i), we need only to show that

(A.4) $\qquad \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \dot{g}(V_t, \theta_1)[\Delta(V_t) + e_t] \xrightarrow{d} N(0, \Sigma_e + \Sigma_\Delta).$

Note that

$$\text{Var}\left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \dot{g}(V_t, \theta_1)[\Delta(V_t) + e_t] \right\}$$
$$= \frac{1}{n} E \left\{ \sum_{t=1}^{n} \dot{g}(V_t, \theta_1)[\Delta(V_t) + e_t] \right\}^2$$
$$= \Sigma_e + \Sigma_\Delta.$$

Then, by the central limit theorem for i.i.d. sequences, (A.4) follows and thus the proof of Theorem 2.2(i) is completed. Theorem 2.2(ii) follows by (2.9) and the standard arguments for local linear estimators. $\qquad\square$

Proof of Theorem 2.3. By a standard argument, we have

(A.5) $\quad \widehat{\theta}_1 - \theta_1$
$$= \left[\sum_{t=1}^{n} \dot{g}(V_t, \theta_1)\dot{g}^{\tau}(V_t, \theta_1) \right]^{-1}$$
$$\times \sum_{t=1}^{n} \dot{g}(V_t, \theta_1)\Delta_n(V_t) \, (1 + o_P(1))$$
$$+ \left[\sum_{t=1}^{n} \dot{g}(V_t, \theta_1)\dot{g}^{\tau}(V_t, \theta_1) \right]^{-1} \sum_{t=1}^{n} \dot{g}(V_t, \theta_1)e_t.$$

By (A.3) and the central limit theorem for i.i.d. sequences, we have

Estimation in semiparametric time series regression 249

$$(A.6) \quad \sqrt{n}\left[\sum_{t=1}^{n}\dot{g}(V_t,\theta_1)\dot{g}^{\tau}(V_t,\theta_1)\right]^{-1}\sum_{t=1}^{n}\dot{g}(V_t,\theta_1)e_t$$

$$\xrightarrow{d} N(0,\Sigma^{-1}(\theta_1)\Sigma_e\Sigma^{-1}(\theta_1)).$$

Meanwhile, as $\{V_t\}$ is a sequence of independent variables, we have

$$(A.7) \quad E\left\|\sum_{t=1}^{n}\{\dot{g}(V_t,\theta_1)\Delta_n(V_t) - E[\dot{g}(V_t,\theta_1)\Delta_n(V_t)]\}\right\|^2$$

$$= O\left(\sum_{t=1}^{n}E\|\dot{g}(V_t,\theta_1)\Delta_n(V_t)\|^2\right).$$

By A5, we have as $n \to \infty$

$$(A.8) \quad \sum_{t=1}^{n}E\|\dot{g}(V_t,\theta_1)\Delta_n(V_t)\|^2$$

$$= n\int_0^1\left[\int\|\dot{g}(v,\theta_1)\Delta_n(v)\|^2 p_u(v-H(r))dv\right]dr$$

$$= O(n\delta_{1n}^2).$$

By (A.7) and (A.8), we also have as $n \to \infty$

$$(A.9) \quad \frac{1}{n}\sum_{t=1}^{n}\dot{g}(V_t,\theta_1)\Delta_n(V_t)$$

$$= \frac{1}{n}\sum_{t=1}^{n}E[\dot{g}(V_t,\theta_1)\Delta_n(V_t)] + O_P(n^{-1/2}\delta_{1n})$$

$$= \int_0^1\left[\int\dot{g}(v,\theta_1)\Delta_n(v)p_u(v-H(r))dv\right]dr$$

$$+ O_P(n^{-1/2}\delta_{1n})$$

$$= O_P(\delta_n + n^{-1/2}\delta_{1n}).$$

By (A.5), (A.6) and (A.9), we prove that Theorem 2.3(i) holds.

We now prove Theorem 2.3(ii). Observe that

$$(A.10)$$

$$\widehat{\Delta}_n(v_0) - \Delta_n(v_0)$$

$$= \sum_{t=1}^{n}W_{nt}(v_0)e_t + \sum_{t=1}^{n}W_{nt}(v_0)[\Delta_n(V_t) - \Delta_n(v_0)]$$

$$+ \sum_{t=1}^{n}W_{nt}(v_0)[g(V_t,\theta_1) - g(V_t,\widehat{\theta}_1)]$$

$$=: J_{n1}(v_0) + J_{n2}(v_0) + J_{n3}(v_0).$$

By the central limit theorem for i.i.d. sequences, we have, as $n \to \infty$,

$$(A.11) \quad \sqrt{nb}J_{n1}(v_0) \xrightarrow{d} N\left(0, \frac{\sigma^2(v_0)\int K^2(u)du}{f(v_0)}\right),$$

where $\sigma^2(v_0) = \int_0^1\sigma^2(v_0,r)dr$ and $f(v_0) = \int_0^1 p_u(v_0 - H(r))dr$.

By Taylor expansion, we have

$$(A.12) \quad J_{n2}(v_0) = \frac{1+o_P(1)}{2}b^2\Delta_n''(v_0)\int u^2 K(u)du.$$

Meanwhile, a straightforward derivation implies that as $n \to \infty$

$$(A.13) \quad \sum_{t=1}^{n}W_{nt}(v_0)\dot{g}(V_t,\theta_1) = \dot{g}(v_0,\theta_1) + o_P(1),$$

which, along with the conclusion of Theorem 2.3(i), gives

$$(A.14) \quad J_{n3}(v_0) = O_P(n^{-1/2} + \delta_n).$$

The proof of Theorem 2.3(ii) therefore follows from (A.10)–(A.14). □

Proof of Corollary 2.4. Note that

$$(A.15)$$

$$\widehat{\theta}_1 - \theta_1 = \left[\sum_{t=1}^{n}\dot{g}(V_t,\theta_1)\dot{g}^{\tau}(V_t,\theta_1)\right]^{-1}$$

$$\times\left\{\sum_{t=1}^{n}\dot{g}(V_t,\theta_1)[\delta_n\Delta(V_t) + e_t]\right\}(1 + o_P(1)).$$

The proof then is similar to that of Theorem 2.2. The main difference lies in the term $\{\sum_{t=1}^{n}\dot{g}(V_t,\theta_1)[\delta_n\ \Delta(V_t) + e_t]\}$. In view of (A.3), to prove Corollary 2.4(i), we need only to show that

$$(A.16) \quad \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\dot{g}(V_t,\theta_1)[\delta_n\ \Delta(V_t) + e_t] \xrightarrow{d} N(0,\Sigma_e).$$

(A.16) then follows by noting that

$$\text{Var}\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\dot{g}(V_t,\theta_1)[\delta_n\ \Delta(V_t) + e_t]\right\}$$

$$= \frac{1}{n}E\left\{\sum_{t=1}^{n}\dot{g}(V_t,\theta_1)[\delta_n\cdot\Delta(V_t) + e_t]\right\}^2 \to \Sigma_e.$$

This completes the proof of Corollary 2.4(i). □

The proof of Corollary 2.4(ii) is similar to that of Theorem 2.3(ii).

ACKNOWLEDGEMENTS

REFERENCES

[1] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modeling and Its Applications*. Chapman & Hall, London. MR1383587

[2] FAN, J., WU, Y. and FENG, Y. (2009). Local quasi-likelihood with a parametric guide. *Annals of Statistics* **37** 4153–4183. MR2572456

[3] FAN, J. and YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York. MR1964455

[4] GAO, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. Chapman & Hall, London. MR2297190

[5] GAO, J. and GIJBELS, I. (2008). Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association* **484** 1584–1594. MR2504206

[6] GAO, J. and HAWTHORNE, K. (2006). Semiparametric estimation and testing of the trend of temperature series. *Econometrics Journal* **9** 332–355. MR2324973

[7] GAO, J. and LIANG, H. (1997). Statistical inference in single-index and partially nonlinear models. *Annals of the Institute of Statistical Mathematics* **49** 493–517. MR1482369

[8] HÄRDLE, W., LIANG, H. and GAO, J. (2000). *Partially Linear Models*. Physica-Verlag, New York. MR1787637

[9] JONES, P. D. (1989). The influence of ENSO on global temperatures. *Climate Monitor* **2** 285–90.

[10] KREISS, J. P., NEUMANN, M. H. and YAO, Q. (2008). Bootstrap tests for simple structures in nonparametric times series regression. *Statistics and Its Interface* **1** 367–380. MR2476752

[11] LI, Q. and RACINE, J. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, Princeton. MR2283034

[12] TONG, H. (1990). *Nonlinear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford. MR1079320

[13] WU, C. F. (1981). Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics* **9** 501–513. MR0615427

[14] XIA, Y., TONG, H. and LI, W. K. (1999). On extended partially linear single-index models. *Biometrika* **86** 831–842. MR1741980

[15] ZHENG, X. and BASHER, R. E. (1999). Structural time series models and trend detection in global and regional temperature series. *Journal of Climate* **12** 2347–58.

Jia Chen
Department of Econometrics and Business Statistics
Monash University Caulfield Campus, VIC 3145
Australia
E-mail address: jiachen1981@gmail.com

Jiti Gao
Department of Econometrics and Business Statistics
Monash University Caulfield Campus, VIC 3145
Australia
E-mail address: jiti.gao@monash.edu

Degui Li
Department of Econometrics and Business Statistics
Monash University Caulfield Campus, VIC 3145
Australia
E-mail address: degui.li08@gmail.com